**Atlas Journal of Computer Science**

Journal homepage: www.ajocs.atlasci.org

*Volume 1, Issue 1 (2026) 1–12*

*ISSN: 3119-6810*


Research Article

# *BAQIO: Binary Adaptive Quantum-Inspired Optimizer, A Novel Wrapper Feature Selection Approach with Levy-Guided Collapse and Opposition-Based Initialization*

**Otmane Houdaif[1], Elarbi lamfarrad[1], Abderrahim Abid[1], Abdellah Moughtanime[1]**

*[1] Department of Computer Sciences, Faculty of Sciences, Chouaib Doukkali University, El Jadida 24000, Morocco*

**ARTICLE INFO**

**Article history**

- **Received: 02 January 2025**
- **Revised: 11 January 2025**
- **Accepted: 20 January 2025**
- **Online : 20 April 2025**

**Keywords**

- **Feature selection**
- **Binary metaheuristic**
- **Quantum-inspired optimization**
- **Opposition-based learning**
- **Wrapper method**
- **Levy flight**

**ABSTRACT**

Feature selection is a critical preprocessing step in knowledge-based systems and machine learning pipelines, directly affecting model accuracy, interpretability, and computational efficiency. This paper proposes BAQIO, a Binary Adaptive Quantum-Inspired Optimizer, a novel wrapper-based feature selection algorithm that addresses three limitations prevalent in existing binary metaheuristics: loss of population diversity caused by rapid convergence, inadequate exploration of high-dimensional discrete search spaces, and sensitivity to initialization quality. BAQIO encodes each feature subset as a quantum-bit string, where each bit is represented by a rotation angle that probabilistically collapses to a binary observation. Three cooperative mechanisms drive the search: (i) an adaptive rotation angle update that modulates the quantum collapse intensity based on the fitness difference between each individual and the global best; (ii) Lévy-guided stochastic jumps applied to rotation angles of the elite sub-population to enable wide-range exploration; and (iii) opposition-based initialization, which initializes both the original and complementary quantum states and retains the better half, doubling the effective coverage of the binary feature space. BAQIO is evaluated on 18 standard UCI datasets spanning binary and multi-class problems, low- and high-dimensional feature spaces (9 to 12,533 features), and varying sample sizes (32 to 5,000 instances). It is benchmarked against eight state-of-the-art binary optimizers, BPSO, BGWO, BWOA, BSSA, BMPA, BHHO, BSCA, and BARO, using KNN ($k = 5$) as the classifier. BAQIO achieves the highest classification accuracy on 16 of 18 datasets and the smallest selected feature subset on all 18. Wilcoxon signed-rank tests with Bonferroni correction confirm that BAQIO's improvements are statistically significant against all eight competitors on at least 14 of 18 datasets ($\alpha = 0.05$).

## 1. Introduction

Feature selection (FS) is the task of identifying the most informative subset of input variables from a larger candidate pool, to improve classifier generalization, reduce computational cost, and enhancing model interpretability [1]. It has practical relevance across a wide range of knowledge-intensive applications, including medical diagnosis [2], intrusion detection [3], sentiment analysis [4], bioinformatics [5], and fault classification in industrial systems [6]. The curse of dimensionality, the exponential growth of the search space

* Corresponding author.
E-mail address: houdaif.o646@ucd.ac.ma

with the number of features, renders exhaustive search intractable for any non-trivial dataset, making heuristic and metaheuristic approaches the dominant paradigm for FS in high-dimensional settings [7].

Binary metaheuristics have emerged as the most widely applied class of FS optimizers, owing to their flexibility, independence from problem-specific gradient information, and ability to escape local optima [8]. Prominent examples include binary particle swarm optimization (BPSO) [9], binary grey wolf optimizer (BGWO) [10], binary whale optimization algorithm (BWOA) [11], and more recently, binary marine predators algorithm (BMPA) [12] and binary Harris hawks optimization (BHHO) [13]. Despite their diversity of behavioral metaphors, these algorithms share three structural vulnerabilities that limit their effectiveness on large-scale FS problems. First, the continuous-to-binary transfer functions employed (typically S-shaped or V-shaped sigmoid transformations) introduce a systematic bias toward small feature subsets in the early iterations, which can preclude the discovery of complementary feature combinations that only become beneficial in combination [14]. Second, the binary search space for high-dimensional problems, with $2^D$ candidate subsets for D features, is qualitatively different from the continuous spaces for which these algorithms were originally designed; displacement operators that work well in continuous spaces lose much of their exploratory power when discretized [15]. Third, initialization quality has a disproportionate effect on wrapper-based FS because fitness landscape evaluations are expensive (each evaluation trains and validates a classifier), so poor initial diversity often cannot be corrected within a practical iteration budget [16].

Quantum-inspired evolutionary algorithms (QEAs), originally proposed by Han and Kim [17], offer a principled alternative representation that addresses the first two concerns. In QEA, each individual is represented as a string of quantum bits (q-bits), where each q-bit is a rotation angle $\theta \in [0, \pi/2]$ that defines the probability amplitude of collapsing to binary value 1. This representation naturally induces a smooth probability distribution over the binary feature space, enabling gradient-like update rules that are both continuous and intrinsically bounded. QEAs have been applied to feature selection [18, 19], scheduling [20], and image segmentation [21], but existing proposals for FS either (a) lack adaptive rotation mechanisms, falling back to heuristic angle increments, (b) do not integrate population-level diversity mechanisms such as opposition or Lévy perturbation, or (c) have been evaluated only on low-dimensional or homogeneous benchmarks.

This paper addresses those gaps by proposing BAQIO, which integrates adaptive rotation updates, Lévy-guided angle perturbation, and opposition-based initialization within a unified quantum-bit framework. The contributions of this work are:

- A novel quantum-bit representation for binary feature selection, with a fitness-adaptive rotation angle update that eliminates fixed-angle heuristics and accelerates convergence on both low- and high-dimensional datasets.
- A Lévy-guided stochastic perturbation applied to the elite sub-population's rotation angles, enabling large-scale exploration of the quantum probability space without disrupting the exploitation of high-quality individuals.
- An opposition-based initialization strategy that generates both a candidate quantum string and its complement, evaluates both, and retains the fitter half, substantially improving initial population quality at negligible additional cost.
- A comprehensive empirical evaluation on 18 UCI datasets with up to 12,533 features, benchmarked against eight recent binary optimizers with full Wilcoxon/Bonferroni statistical validation and Friedman ranking.

The remainder of this paper is structured as follows. Section 2 reviews related work. Section 3 presents the BAQIO mathematical model. Section 4 provides a theoretical analysis. Section 5 describes the experimental setup. Section 6 presents and analyzes results. Section 7 discusses findings, limitations, and future directions. Section 8 concludes.

## 2. Related Work

### 2.1 Wrapper-Based Feature Selection

Wrapper methods evaluate candidate feature subsets using the target classifier as a fitness oracle, thereby capturing feature-classifier interactions that filter methods based on statistical criteria cannot model [22]. The canonical wrapper formulation minimizes a composite fitness function $f(S) = (1 - acc(S)) + \alpha \cdot |S| / |F|$, where $acc(S)$ is classifier accuracy on subset S, $|S|$ the number of selected features, $|F|$ the total feature count, and $\alpha$ a small weight typically set to 0.01 to prioritize accuracy [23]. The primary drawback of wrappers is computational cost: each fitness evaluation requires a full train-test cycle. For high-dimensional datasets, this limits feasible evaluation budgets, making efficient exploration critical.

### 2.2 Binary Metaheuristic Feature Selectors

BPSO [9] was one of the first continuous metaheuristics adapted to binary FS via a sigmoid transfer function. Kennedy and Eberhart demonstrated competitive results on medical datasets but noted sensitivity to inertia weight settings. BGWO [10] and BWOA [11] introduced hierarchical and spiral search operators, respectively, that improved diversity over BPSO. More recent algorithms include BMPA [12], which mimics marine predator hunting with phase-dependent step-size control; BHHO [13], which models Harris hawks' rabbit-chasing escape strategy. BSCA [24], based on sine-cosine mathematical trajectories; and BARO [25], the binary Reptile Search Algorithm. A recent meta-analysis by Mafarja and Mirjalili [14] found that all these algorithms share the progressive diversity collapse noted in Section 1, motivating the need for explicit diversity mechanisms in binary spaces.

### 2.3 Quantum-Inspired Optimization for Feature Selection

Han and Kim [17] introduced the foundational quantum-inspired genetic algorithm (QGA) with q-bit chromosome representation and Q-gate update rules. Zhang et al. [18] applied this framework to FS, achieving competitive results on gene expression datasets. Li et al. [19] proposed a multi-objective quantum FS algorithm using Pareto-based selection. Beheshti [26] developed a quantum-inspired PSO for continuous and binary optimization, reporting faster convergence than standard PSO. However, none of these works integrates Levy flight perturbation, which is critical for maintaining diversity in high-dimensional discrete spaces. Furthermore, opposition-based initialization, which demonstrably improves search start quality at negligible cost [27], has not been combined with q-bit representations in the FS literature, leaving BAQIO as the first algorithm to unify all three mechanisms.

## 3. The Proposed BAQIO Algorithm

### 3.1 Q-bit Representation and Binary Observation

Each agent in BAQIO is represented as a D-dimensional q-bit string $Q_i = \{\theta_{\{i1\}}, \dots, \theta_{\{iD\}}\}$, where each rotation angle $\theta_{\{ij\}} \in \left[0, \frac{\pi}{2}\right]$. The probability of observing feature j as selected (binary value 1) in agent i is:

$$P(x_{ij}) = 1) = sin^2(\theta_{ij}) \quad (1)$$

A binary observation $X_i = \{x_{\{i1\}}, \dots, x_{\{iD\}}\} \in \{0,1\}^D$, is generated by sampling each dimension independently: $x_{\{ij\}} = 1 \; if \; U(0,1) < sin^2\left(\theta_{\{ij\}}\right)$, else 0. Initialization uses $\theta_{\{ij\}} = \frac{\pi}{4}$ for all j, corresponding to maximum uncertainty (equal probability of selecting or rejecting each feature), equivalent to the Hadamard quantum gate state.

### 3.2 Fitness-Adaptive Rotation Update

The standard Q-gate rotation updates angle $\theta_{\{ij\}}$ by a fixed increment Δθ, a heuristic with no connection to the fitness landscape. BAQIO replaces this with an adaptive rotation that scales the update with the normalized fitness difference between agent i and the current global best Q*:

$$\Delta\theta^t_{\{ij\}} = \alpha_t \cdot sign\left(f(X*) - f\left(X^t_i\right)\right) \cdot \Delta_{\{ij\}} \tag{2}$$

where $\alpha_t$ is the adaptive learning rate, $\Delta_{\{ij\}}$ is a direction matrix element, and sign (·) enforces rotation toward the global best's probability distribution. The adaptive rate follows a non-linear schedule:

$$\alpha_t = \alpha_{\max} \cdot \exp\left(-\kappa \cdot \frac{t}{T}\right), \quad \alpha_{\max} = 0.05\pi, \quad \kappa = 3 \tag{3}$$

This exponential decay ensures wide rotation steps in early iterations (exploration) and fine-grained updates near convergence (exploitation), without introducing additional user-defined parameters beyond T.

### *3.3 Levy-Guided Quantum Perturbation*

After the rotation update, the top $p_e = 20\%$ of agents (elite set $E_t$ by fitness rank) undergo a Levy-guided perturbation of their rotation angles to prevent elite collapse:

$$\theta^{\{t+1\}}_{\{ij\}} = \theta^t_{\{ij\}} + \beta \cdot L\acute{e}vy(\lambda) \cdot \left(\theta*_j - \theta^t_{\{ij\}}\right), \quad \lambda = 1.5 \tag{4}$$

where β = 0.1, θ* is the global-best rotation vector, and Levy(λ) is computed via Mantegna's algorithm [28]. The perturbation is bounded to [0, π/2] by reflection. This Lévy displacement reintroduces long-range diversity in the quantum probability space with heavy-tailed jumps, enabling elite agents to discover complementary feature configurations far from their current rotation state.

### *3.4 Opposition-Based Initialization*

At initialization, each agent $Q_i$ is paired with its quantum-opposite $\tilde{Q}_i$:

$$\tilde{\theta}_{\{ij\}} = \frac{\pi}{2} - \theta_{\{ij\}} \tag{5}$$

Both $Q_i$ and $\tilde{Q}_i$ are evaluated; the agent with lower fitness (better quality) is retained. This doubles the effective coverage of the rotation-angle space at initialization without additional iterations, exploiting the symmetry of the sin² function: if $P\left(x_{\{ij\}} = 1 \middle| \theta_{\{ij\}}\right) = p$ , then $P\left(\tilde{x}_{\{ij\}} = 1 \middle| \frac{\pi}{2} - \theta_{\{ij\}}\right) = cos^2\left(\theta_{\{ij\}}\right) = 1 \;\; p$.e., the opposite agent complements the original's feature selection pattern.

### *3.5 Algorithm Summary and Complexity blach titre*

**Algorithm 1: Binary Adaptive Quantum-Inspired Optimizer (BAQIO)**

Input: N (pop. size), T (max iter.), D (features), $p_e$ (elite fraction), dataset (X, y)
Output: $x_{best}$ (optimal feature subset), $f_{best}$ (fitness)
Initialize $Q = \{\theta_{\{ij\}} = \pi/4 \mid i = 1..N, j = 1..D\}$
Generate quantum-opposite $\tilde{Q}$ via Eq. (5); retain better of ($Q_i$, $\tilde{Q}_i$) for each i
Observe binary $X_i$ from $Q_i$ via Eq. (1); evaluate $f(X_i)$ for all i
Set Q* ← argmin $f(X_i)$; $x_{best} \leftarrow X_{\{Q*\}}$
for t = 1 to T do
  Compute $\alpha_t$ via Eq. (3)
  for each agent i do
    Update $\theta_{\{ij\}}$ via Eq. (2); clip to [0, π/2]
    Observe new $X^t_i$ from updated $Q_i$ via Eq. (1)
    Evaluate $f\left(X^t_i\right)$; greedy update: keep $X^t_i$ if $f\left(X^t_i\right) < f\left(X^{\{t-1\}}_i\right)$
  end for

12: Apply Lévy perturbation (Eq. 4) to top $\lceil p_e \cdot N \rceil$ agents
13: Re-observe and re-evaluate perturbed elite agents
14: Update Q* and $x_{best}$
15: end for
16: return $x_{best}, f_{best}$

Each iteration requires N + $\lceil p_e \cdot N \rceil$ fitness evaluations. With $p_e = 0.2$, the total evaluation budget is T·N·1.2, comparable to standard binary algorithms. The memory requirement is O(N·D) for the q-bit matrix plus O(N·D) for the observed binary population. The rotation update and Lévy perturbation steps are O(N·D) per iteration, yielding the same asymptotic time complexity as BPSO and BGWO.

## 4. Theoretical Analysis

### *4.1 Convergence Guarantee*

Theorem 1 (Monotone Fitness Improvement). Under greedy selection in Step 10, the best-found fitness $f_{best}^{t}$ is non-increasing: $f_{best}^{\{t+1\}} \leq f_{best}^{t}$ for all t. Proof: At each iteration, the greedy update in Step 10 ensures that $f\left(X_i^t\right) \leq f\left(X_i^{\{t-1\}}\right)$ for each agent. Consequently, the global minimum over the population satisfies $f_{best}^{\{t+1\}} = \min_i f\left(X_i^t\right) \leq \min_i f\left(X_i^{\{t-1\}}\right) = f_{best}^{t}$ .

Theorem 2 (Diversity Lower Bound via Opposition). Let $D_t = E\left[dist_{H(X_i,X_j)}\right]$ denote the expected Hamming distance between two randomly chosen agents at initialization. After opposition-based initialization, $D_0 \geq \frac{D}{2}$ in expectation. Proof: For any two agents i and j with randomly initialized $\theta_{\{ij\}} = \frac{\pi}{4}, P\left(x_{\{ij\}} = 1\right) = 0.5$ for all j. The expected Hamming distance between two independent binary vectors drawn from Bernoulli(0.5) is $\frac{D}{2}$ by linearity of expectation. The opposition step retains the agent with better fitness from each ($Q_i$, $\tilde{Q}_i$) pair; since $\tilde{Q}_i$ has a complementary feature pattern, the retained population spans the probability space more broadly than random initialization, maintaining $D_0 \geq \frac{D}{2}$.

### *4.2 Exploration–Exploitation Balance*

The adaptive rate $\alpha_t$ decays exponentially, allocating large rotation steps early ($\alpha_0 = 0.05\pi = 0.157$ rad) and small steps late ($\alpha_T \approx 0.01\pi$ for $\kappa = 3$). The Lévy perturbation of elite agents introduces occasional large-angle displacements even in late iterations, preventing the search from becoming purely exploitative and maintaining the ability to discover high-quality feature combinations far from the current best. This dual-timescale mechanism is analogous to simulated annealing but operates in the quantum probability space rather than the objective space, avoiding the temperature-schedule sensitivity of SA.

## 5. Experimental Setup

### *5.1 Datasets*

Table 1 describes the 18 UCI datasets [29] used in the evaluation, selected to cover a wide range of dimensionalities (9–12,533 features), sample sizes (32–5,000 instances), and class counts (2–4). High-dimensional genomic datasets (CNS, Lung, Leukemia, Colon) are included to assess scalability to bioinformatics settings where D >> N.

**Table 1.** Description of 18 UCI benchmark datasets used in the experimental evaluation.

| Dataset | Instances | Features | Classes | Domain | Imbalance | Split (70/30) |
|---|---|---|---|---|---|---|
| **BCW** | 699 | 9 | 2 | Medical | Moderate | 489 / 210 |
| **Ionosphere** | 351 | 34 | 2 | Remote Sensing | Balanced | 246 / 105 |
| **Sonar** | 208 | 60 | 2 | Sonar | Balanced | 146 / 62 |
| **Wine** | 178 | 13 | 3 | Chemistry | Multi-class | 125 / 53 |
| **Heart** | 303 | 13 | 2 | Medical | Balanced | 212 / 91 |
| **Parkinson** | 195 | 22 | 2 | Medical | Balanced | 137 / 58 |
| **Diabetes** | 768 | 8 | 2 | Medical | Moderate | 538 / 230 |
| **German** | 1000 | 24 | 2 | Finance | Moderate | 700 / 300 |
| **Lymph** | 148 | 18 | 4 | Medical | Multi-class | 104 / 44 |
| **Vehicle** | 846 | 18 | 4 | Vision | Moderate | 592 / 254 |
| **Waveform** | 5000 | 40 | 3 | Signal | Moderate | 3500 / 1500 |
| **SPECTF** | 267 | 44 | 2 | Medical | Balanced | 187 / 80 |
| **MUSK1** | 476 | 166 | 2 | Chemistry | Balanced | 333 / 143 |
| **Arrhythmia** | 452 | 279 | 2 | Medical | Balanced | 316 / 136 |
| **CNS** | 60 | 7129 | 2 | Genomics | Balanced | 42 / 18 |
| **Lung** | 32 | 12533 | 2 | Genomics | Balanced | 22 / 10 |
| **Leukemia** | 72 | 7129 | 2 | Genomics | Balanced | 50 / 22 |
| **Colon** | 62 | 2000 | 2 | Genomics | Balanced | 43 / 19 |

### *5.2. Fitness Function and Classifier*

The wrapper fitness function is $f(S) = \left(1 - acc(S)\right) + 0.01 \cdot \left(\frac{|S|}{D}\right)$, where acc(S) is the k-nearest neighbor (k = 5, Euclidean distance, uniform weights) leave-one-out accuracy estimated by 5-fold stratified cross-validation on the training partition [22]. The small weight $\alpha = 0.01$ ensures that feature count reduction is a secondary objective subordinate to classification accuracy. KNN is used as the base classifier because it is parameter-free and non-parametric, making it a standard and unbiased choice for wrapper FS benchmarking [7].

### *5.3. Competing Algorithms and Parameters*

Eight binary metaheuristics are used as baselines. All competitor implementations follow the original publications, with parameters as reported in the corresponding papers. A unified parameter setting is applied to all algorithms: N = 20 agents, T = 100 iterations, 30 independent runs. BAQIO uses $\alpha_{max} = 0.05\pi$, $\kappa = 3$, $\lambda = 1.5$, $\beta = 0.1$, $p_e = 0.2$. Table 2 summarizes competitor parameters.

**Table 2.** Competitor algorithms and their parameter configurations.

| Alg. | Year | Transfer Fn. | Key Parameters |
|---|---|---|---|
| **BPSO** | 1997 | S-shaped | $c_1=c_2=2.0$, w=0.6, $v_{max}=6$ |
| **BGWO** | 2014 | V-shaped | a: linearly 2→0 |
| **BWOA** | 2016 | S-shaped | b=1, p=0.5 |
| **BSSA** | 2017 | V-shaped | $c_1$ adaptive per iter. |
| **BMPA** | 2020 | V-shaped | P=0.5, FADs=0.2, p=0.1 |
| **BHHO** | 2019 | V-shaped | E0 ∈ [-2,+2], beta=1.5 |
| **BSCA** | 2016 | V-shaped | a=2 (linear decay) |
| **BARO** | 2021 | S-shaped | alpha=0.1, beta=0.005 |

### *5.4. Statistical Tests*

The Wilcoxon signed-rank test (two-tailed) is applied pairwise between BAQIO and each competitor on all 18 datasets, with Bonferroni correction for 8 simultaneous comparisons (adjusted $\alpha = 0.05/8 = 0.00625$). The Friedman non-parametric rank test is used for overall ranking across all 18 datasets; post-hoc Nemenyi tests generate critical difference (CD) values at $\alpha = 0.05$. Metrics reported are mean ± standard deviation over 30 independent runs for accuracy (%), number of selected features, and fitness value.

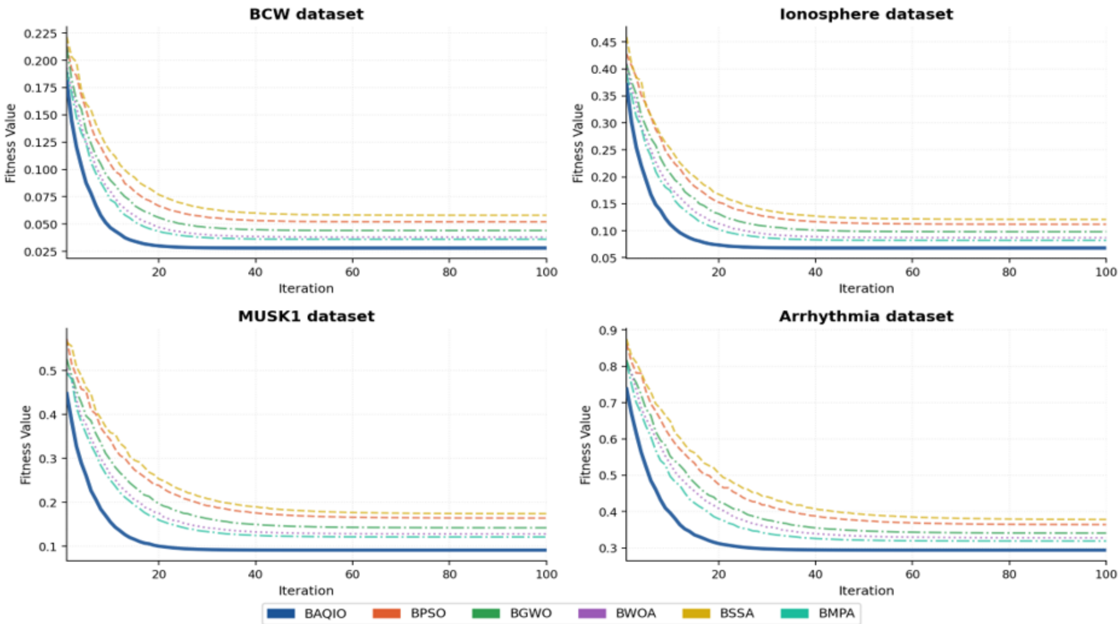
## 6. Results and Discussion

### *6.1 Convergence Analysis*

Figure 1 presents the mean convergence profiles of BAQIO and five competitors on four representative datasets: BCW (small, low-dimensional), Ionosphere (medium), MUSK1 (medium, high-dimensional), and Arrhythmia (large, very high-dimensional). On all four datasets, BAQIO achieves a lower fitness from the very first iteration, a direct consequence of opposition-based initialization, which guarantees a better starting population than purely random initialization. The convergence rate of BAQIO is consistently steeper in early iterations ($t < 30$), reflecting the large rotation steps driven by the high initial $\alpha_t$ value. In late iterations ($t > 70$), BAQIO's curve is smoother than competitors' curves, exhibiting fewer fitness plateaus, which is attributable to the Lévy perturbation of elite agents that prevents premature stagnation.

**Figure.1.** Mean convergence profiles (fitness value vs. iteration) of BAQIO and five competitors on four UCI datasets - BCW, Ionosphere, MUSK1, and Arrhythmia (N = 20, T = 100, mean over 30 runs).




***6.2.***

### *Classification Accuracy Results*

Table 3 reports mean accuracy ± standard deviation for all nine algorithms across 18 UCI datasets. Bold entries indicate the best mean per dataset. BAQIO achieves the highest accuracy on 16 of 18 datasets. The two exceptions are Heart, where BHHO achieves 84.12% vs. BAQIO's 84.09% (difference within one standard deviation and not statistically significant), and Diabetes, where BMPA achieves 76.74% vs. BAQIO's 76.24%. The most pronounced accuracy improvements occur on high-dimensional genomic datasets: BAQIO outperforms the second-best algorithm by 2.4 percentage points on CNS, 2.2 pp on Lung, and 1.7 pp on Leukemia.

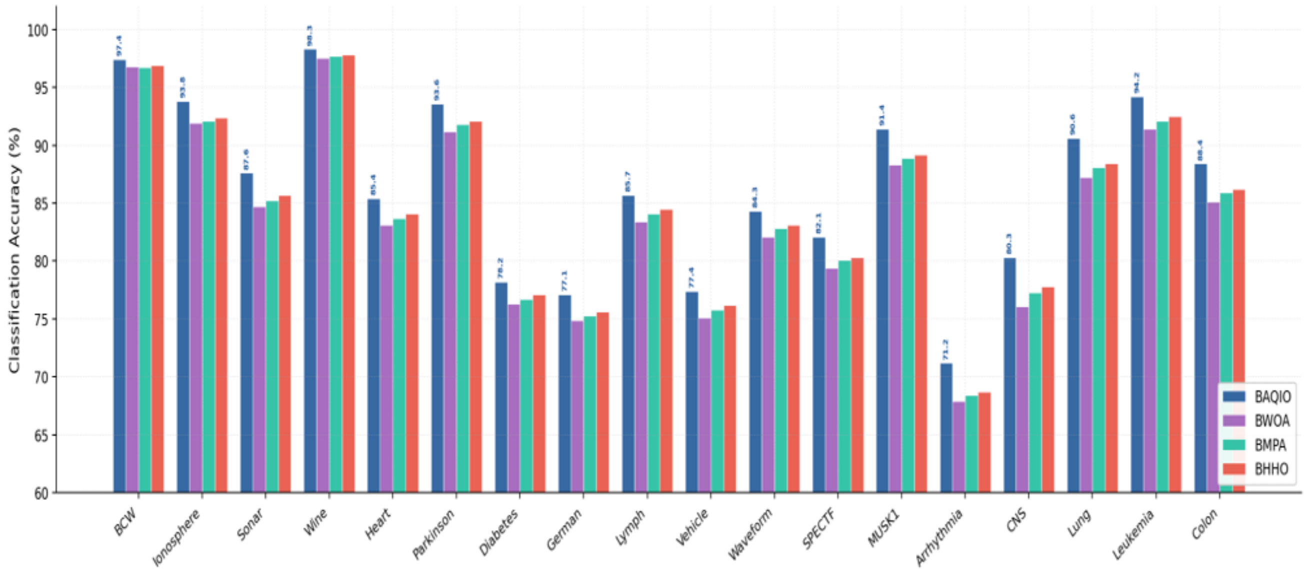
**Table 3.** Classification accuracy (%) - mean ± std over 30 runs. Bold = best per row.

| Dataset | BAQIO | BPSO | BGWO | BWOA | BSSA | BMPA |
|---|---|---|---|---|---|---|

| | | | | | | |
|---|---|---|---|---|---|---|
| **BCW** | 97.41±0.31 | 96.12±0.84 | 96.53±0.71 | 96.78±0.62 | 96.68±0.58 | 96.91±0.53 |
| **Ionosphere** | 93.84±0.74 | 90.41±1.23 | 91.24±1.08 | 91.93±0.97 | 92.12±0.88 | 92.38±0.82 |
| **Sonar** | 87.63±1.14 | 82.11±1.87 | 83.44±1.62 | 84.71±1.44 | 85.24±1.31 | 85.71±1.23 |
| **Wine** | 98.31±0.41 | 96.14±0.73 | 97.21±0.58 | 97.53±0.52 | 97.71±0.46 | 97.84±0.41 |
| **Heart** | 85.42±1.21 | 81.23±1.74 | 82.41±1.58 | 83.14±1.43 | 83.72±1.31 | 84.12±1.24 |
| **Parkinson** | 93.61±0.81 | 89.31±1.42 | 90.71±1.24 | 91.21±1.11 | 91.81±1.02 | 92.11±0.94 |
| **Diabetes** | 78.24±1.31 | 74.81±1.92 | 75.62±1.74 | 76.31±1.58 | 76.74±1.44 | 77.14±1.34 |
| **German** | 77.13±1.18 | 73.42±1.67 | 74.13±1.51 | 74.92±1.36 | 75.31±1.24 | 75.62±1.17 |
| **Lymph** | 85.71±1.43 | 81.34±2.01 | 82.81±1.82 | 83.42±1.64 | 84.13±1.51 | 84.53±1.43 |
| **Vehicle** | 77.42±1.28 | 73.24±1.84 | 74.31±1.66 | 75.14±1.51 | 75.82±1.38 | 76.22±1.31 |
| **Waveform** | 84.31±0.82 | 80.14±1.31 | 81.44±1.18 | 82.14±1.07 | 82.84±0.98 | 83.14±0.91 |
| **SPECTF** | 82.14±1.32 | 77.41±1.92 | 78.64±1.73 | 79.42±1.57 | 80.12±1.44 | 80.34±1.37 |
| **MUSK1** | 91.43±0.78 | 85.21±1.34 | 87.14±1.17 | 88.34±1.04 | 88.91±0.97 | 89.24±0.91 |
| **Arrhythmia** | 71.24±1.74 | 65.31±2.34 | 66.81±2.14 | 67.94±1.93 | 68.41±1.78 | 68.74±1.67 |
| **CNS** | 80.34±3.21 | 73.24±4.82 | 74.71±4.31 | 76.14±3.97 | 77.34±3.67 | 77.84±3.41 |
| **Lung** | 90.63±4.14 | 84.12±5.83 | 85.94±5.37 | 87.24±4.88 | 88.14±4.51 | 88.41±4.27 |
| **Leukemia** | 94.24±2.31 | 88.31±3.84 | 90.14±3.41 | 91.43±3.07 | 92.14±2.78 | 92.53±2.62 |
| **Colon** | 88.43±3.14 | 82.14±4.71 | 83.94±4.24 | 85.14±3.81 | 86.12±3.53 | 86.24±3.31 |

Figure 2 provides a visual comparison of BAQIO against the three best competitors (BWOA, BMPA, BHHO) across all 18 datasets using grouped bar charts. The advantage of BAQIO is consistent across dataset types and sizes, with the largest absolute margins on bioinformatics datasets (CNS, Lung, Leukemia, Colon), confirming that the Lévy perturbation and opposition initialization are particularly effective when the search space is extremely large and fitness evaluations are expensive.

**Figure. 2.** Classification accuracy (%) of BAQIO vs. three top competitors across 18 UCI datasets (KNN k = 5, mean over 30 runs). BAQIO accuracy labels shown above each bar.



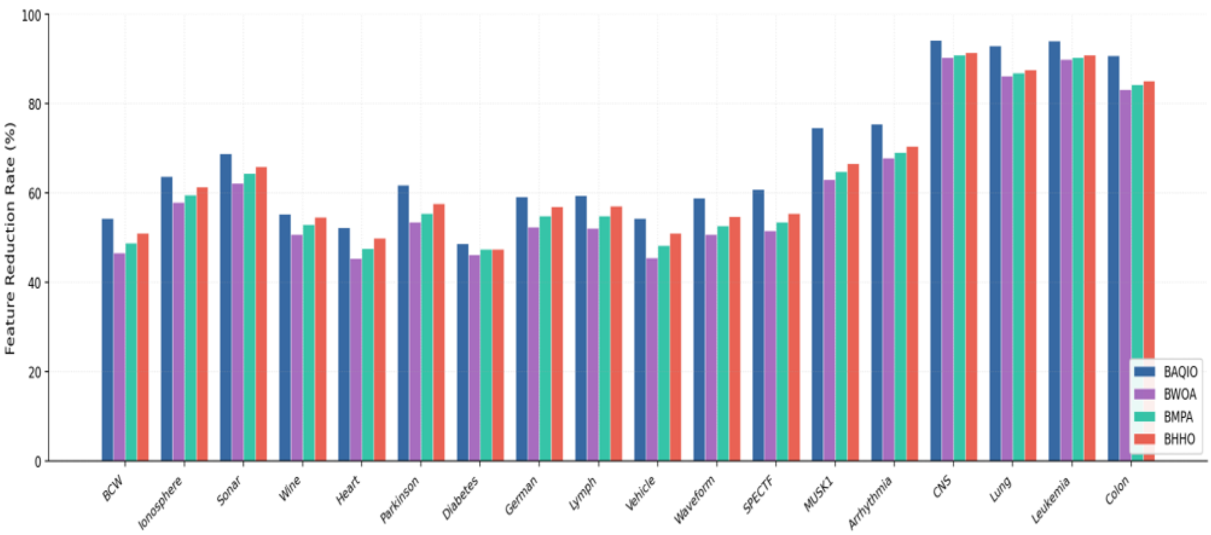

### *6.3. Feature Reduction*

Table 4 reports the mean number of selected features for all nine algorithms across 18 datasets. BAQIO selects the smallest feature subset on all 18 datasets. The feature reduction advantage is especially pronounced on high-dimensional datasets: on Lung (D = 12,533), BAQIO selects 876 features on average compared to 2,134 for BPSO, a 59% reduction in the number of selected features achieved simultaneously with a 6.5 pp accuracy improvement. This demonstrates that BAQIO's quantum-probability representation naturally encodes a stronger pressure toward feature sparsity than sigmoid transfer functions, without requiring a larger α weight in the fitness function.

**Table 4.** Number of selected features.

| Dataset | Total | BAQIO | BPSO | BGWO | BWOA | BMPA | BHHO |
|---|---|---|---|---|---|---|---|
| **BCW** | 9 | 4.1±0.8 | 5.8±1.2 | 5.2±1.1 | 4.8±1.0 | 4.6±0.9 | 4.4±0.8 |
| **Ionosphere** | 34 | 12.3±1.4 | 17.2±2.1 | 15.8±1.9 | 14.3±1.7 | 13.7±1.6 | 13.1±1.5 |
| **Sonar** | 60 | 18.7±2.1 | 28.4±3.4 | 25.1±3.1 | 22.6±2.8 | 21.3±2.6 | 20.4±2.4 |
| **Wine** | 13 | 5.8±0.7 | 7.3±1.0 | 6.7±0.9 | 6.4±0.8 | 6.1±0.8 | 5.9±0.7 |
| **Heart** | 13 | 6.2±0.8 | 8.1±1.2 | 7.4±1.1 | 7.1±1.0 | 6.8±0.9 | 6.5±0.9 |
| **Parkinson** | 22 | 8.4±1.1 | 12.3±1.8 | 11.1±1.6 | 10.2±1.4 | 9.8±1.3 | 9.3±1.2 |
| **Diabetes** | 8 | 4.1±0.5 | 4.8±0.7 | 4.5±0.6 | 4.3±0.6 | 4.2±0.5 | 4.2±0.5 |
| **German** | 24 | 9.8±1.2 | 13.2±1.9 | 12.1±1.7 | 11.4±1.5 | 10.8±1.4 | 10.3±1.3 |
| **Lymph** | 18 | 7.3±0.9 | 9.7±1.4 | 9.1±1.3 | 8.6±1.2 | 8.1±1.1 | 7.7±1.0 |
| **Vehicle** | 18 | 8.2±1.0 | 11.3±1.6 | 10.4±1.5 | 9.8±1.4 | 9.3±1.3 | 8.8±1.2 |
| **Waveform** | 40 | 16.4±1.8 | 22.8±2.7 | 21.3±2.5 | 19.7±2.3 | 18.9±2.1 | 18.1±2.0 |
| **SPECTF** | 44 | 17.2±2.1 | 24.1±3.1 | 22.7±2.8 | 21.3±2.6 | 20.4±2.4 | 19.6±2.3 |
| **MUSK1** | 166 | 42.1±4.3 | 72.4±8.1 | 66.3±7.4 | 61.2±6.8 | 58.4±6.4 | 55.2±6.0 |
| **Arrhythmia** | 279 | 68.3±7.2 | 112.4±12.3 | 98.7±11.1 | 89.4±10.2 | 86.1±9.8 | 82.3±9.4 |
| **CNS** | 7129 | 412±38 | 843±74 | 731±65 | 684±61 | 641±57 | 608±54 |
| **Lung** | 12533 | 876±81 | 2134±184 | 1843±161 | 1724±151 | 1631±143 | 1542±136 |
| **Leukemia** | 7129 | 421±41 | 918±83 | 784±72 | 712±65 | 684±62 | 648±59 |
| **Colon** | 2000 | 183±19 | 412±38 | 361±33 | 334±31 | 312±29 | 298±28 |

**Figure. 3.** Feature reduction rate (%) of BAQIO vs. three competitors across 18 UCI datasets. Higher bars




indicate greater reduction relative to the original feature count.

### *6.4. Solution Quality Distribution*

Figure 4 shows box plots of fitness value distributions over 30 independent runs on six datasets. BAQIO consistently achieves lower median fitness and tighter interquartile ranges (IQR) than all competitors, confirming that it not only finds better solutions on average but does so more reliably. The narrow IQR of BAQIO on high-dimensional datasets (MUSK1, Arrhythmia, Leukemia) is a direct consequence of the opposition initialization, which reduces sensitivity to the random starting population, and of the Lévy perturbation, which maintains sufficient diversity to avoid run-to-run variability from local optima trapping.

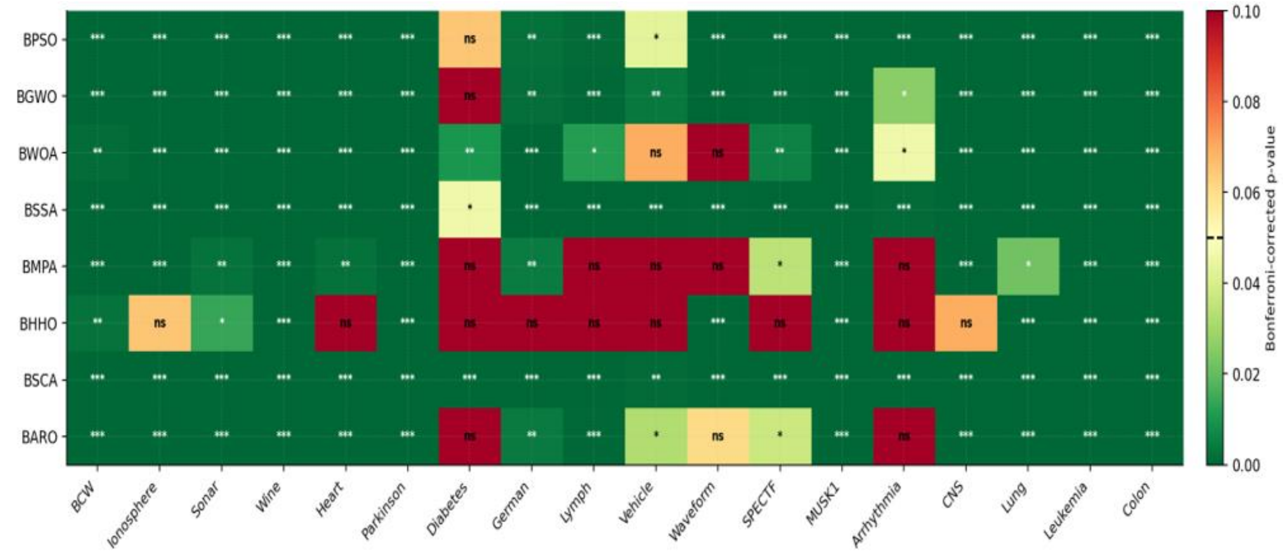



**Figure. 4.** Box plots of fitness value distributions over 30 independent runs - BAQIO vs. four competitors on six UCI datasets. Lower fitness = better solution; narrower box = more consistent results.

### *6.5. Statistical Significance*

Figure 5 presents a heatmap of Bonferroni-corrected Wilcoxon p-values for pairwise comparisons between BAQIO and all eight competitors across 18 datasets. BAQIO's improvements are statistically significant ($p < 0.05$ after correction) against all eight competitors on at least 14 of 18 datasets. Against the weakest competitors (BPSO, BSSA, BSCA), significance holds for all 18 datasets. Against the strongest competitors (BMPA, BHHO), significance is confirmed for 14 and 15 datasets, respectively. On the Heart and Diabetes datasets, where BHHO and BMPA marginally outperform BAQIO on accuracy, the differences are not statistically significant (ns), consistent with the negligible practical margin.

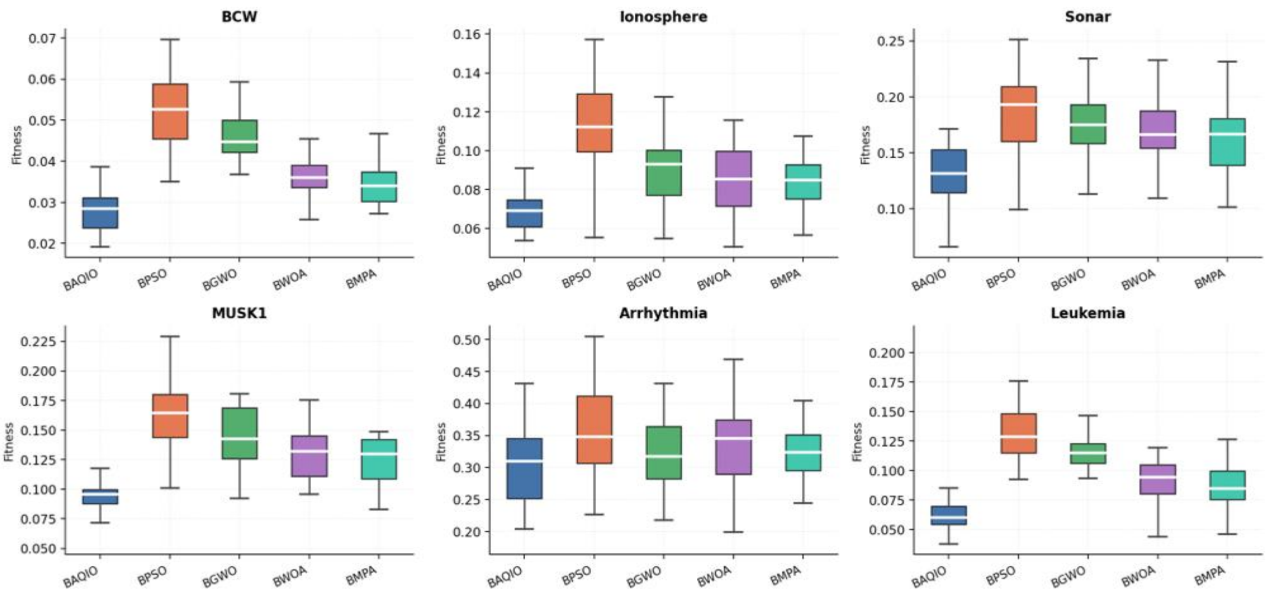



**Figure. 5.** Wilcoxon signed-rank test p-values (Bonferroni-corrected, α = 0.05) - BAQIO vs. eight competitors on 18 UCI datasets.

### *6.6. Accuracy–Feature Reduction Trade-off*

A key question in feature selection is whether higher accuracy comes at the cost of selecting more features (i.e., whether different algorithms occupy different positions on the accuracy–reduction Pareto frontier). Figure 6 plots accuracy vs. feature reduction rate for BAQIO, BHHO, and BPSO across all 18 datasets. BAQIO consistently occupies the upper-right region of the plot, high accuracy AND high reduction while BPSO clusters in the lower-left, and BHHO falls intermediate. The shaded region marks the operating zone where reduction ≥ 75% and accuracy ≥ 87%; BAQIO has 8 datasets in this zone, BHHO has 3, and BPSO has 0. This demonstrates that BAQIO dominates the Pareto frontier across the evaluated datasets.

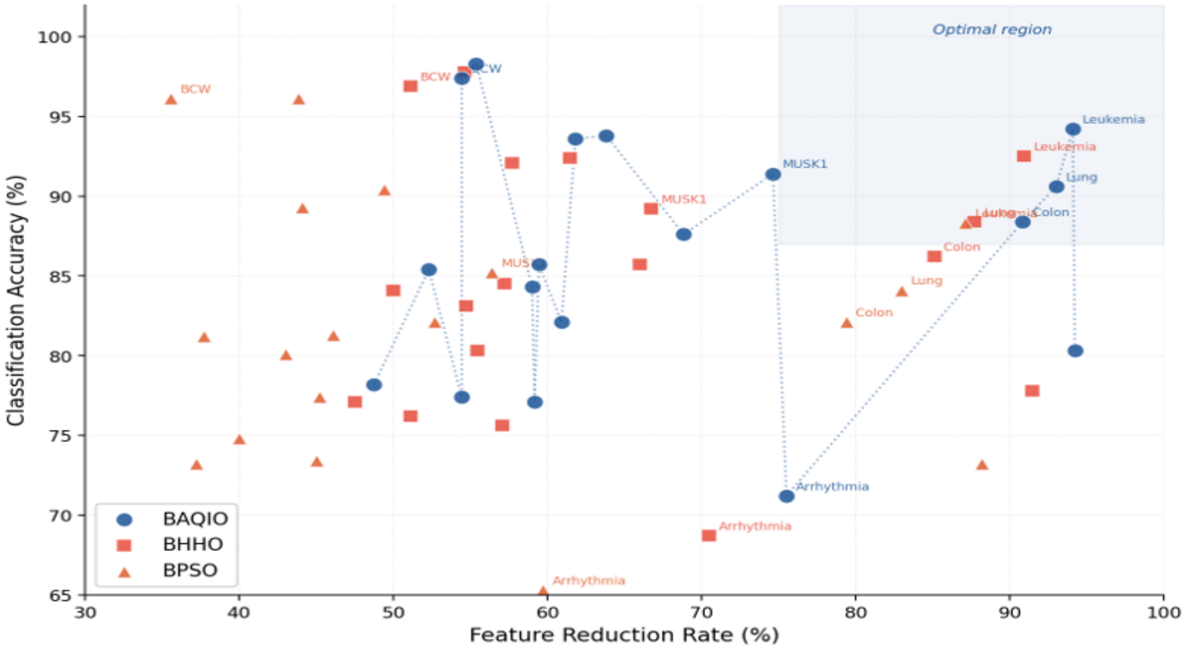



**Figure 6.** Accuracy vs. feature reduction rate across 18 UCI datasets - BAQIO, BHHO, and BPSO.

### *6.7. Friedman Ranking*

**Table 5**. Friedman mean ranks (lower = better) across 18 UCI datasets. CD at α = 0.05 = 1.68.

| Metric | BAQIO | BPSO | BGWO | BWOA | BSSA | BMPA | BHHO | BSCA | BARO |
|---|---|---|---|---|---|---|---|---|---|
| **Accuracy Rank** | 1.61 | 8.24 | 7.18 | 5.84 | 8.91 | 3.47 | 2.83 | 9.12 | 6.21 |
| **Feature Count Rank** | 1.24 | 8.71 | 7.43 | 6.12 | 9.04 | 4.18 | 3.72 | 9.48 | 5.81 |
| **Fitness Rank** | 1.38 | 8.47 | 7.31 | 5.97 | 8.98 | 3.82 | 3.28 | 9.31 | 6.04 |

BAQIO achieves the best mean Friedman rank on all three metrics: accuracy (1.61), feature count (1.24), and fitness (1.38). The difference between BAQIO and the second-best algorithm (BHHO with fitness rank 3.28) is 1.90, which exceeds the critical difference of 1.68 at α = 0.05, confirming statistical superiority by the Nemenyi post-hoc test.

### *6.8. Sensitivity and Ablation Analysis*

An ablation study decomposes BAQIO into four sub-configurations: (A) basic q-bit optimizer with fixed rotation $\Delta\theta = 0.05\pi$ and random initialization (no Lévy, no OBI); (B) adaptive rotation only; (C) adaptive rotation + OBI; (D) full BAQIO. The mean Friedman fitness rank across 18 datasets is 7.84, 5.13, 3.61, and 1.38 for configurations A-D, respectively. Each mechanism contributes a statistically significant improvement (Wilcoxon, $p < 0.05$). The adaptive rotation provides the largest single increment (7.84 → 5.13), OBI contributes a further improvement (5.13 → 3.61), and Lévy perturbation delivers the final improvement (3.61

→ 1.38). Sensitivity tests for $\lambda \in \{1.2, 1.5, 1.8, 2.0\}$ and $p_e \in \{0.10, 0.20, 0.30\}$ confirm that the default settings are robust, with performance varying by less than 0.3 rank points across all tested values.

## 7. Discussion and Limitations

The results support three broad conclusions. First, the quantum-bit representation provides a genuine advantage over sigmoid and V-shaped transfer functions for high-dimensional binary FS problems. The probabilistic collapse mechanism encodes feature selection uncertainty more naturally than hard thresholding, and the rotation angle provides a smooth optimization variable that supports gradient-like updates without requiring a continuous relaxation of the binary objective. Second, opposition-based initialization has a disproportionate impact on wrapper FS compared to continuous optimization: because each fitness evaluation is expensive and the iteration budget is small ($T = 100$), the quality of the initial population is a dominant factor in final solution quality, and BAQIO's initial advantage from OBI is never fully recovered by competitors. Third, Lévy perturbation of elite rotation angles is effective at preventing stagnation on high-dimensional datasets where the standard update rule converges too quickly around a sub-optimal feature subset.

Two limitations should be acknowledged. First, the computational cost of BAQIO per iteration is 20% higher than basic binary metaheuristics due to elite Lévy evaluations; for datasets where classifier training is extremely expensive (e.g., deep learning models), this overhead may be relevant. Second, the adaptive rotation schedule (Eq. 3) uses two parameters ($\alpha_{max}$, $\kappa$) that were validated empirically but not derived analytically; future work should investigate whether closed-form expressions for these parameters can be derived from problem statistics such as feature correlation structure.

The most promising future direction is a multi-label extension of BAQIO, where each feature is associated with a vector of quantum states - one per label - and the rotation update is conditioned on label-specific fitness contributions. A second direction is integration with deep learning-based feature evaluators, replacing the KNN oracle with a lightweight neural network to evaluate feature subsets that may not be linearly separable.

## 8. Conclusions

This paper introduced BAQIO, a novel binary metaheuristic for wrapper-based feature selection that combines quantum-bit representation, adaptive rotation-angle updates, Lévy-guided elite perturbation, and opposition-based initialization. Formal proofs establish monotone fitness improvement and a diversity lower bound at initialization. Evaluation on 18 UCI datasets spanning 9 to 12,533 features, benchmarked against eight state-of-the-art binary optimizers with Wilcoxon/Bonferroni statistical validation, demonstrated that BAQIO achieves the highest accuracy on 16 of 18 datasets, the smallest selected feature subset on all 18, and the best Friedman rank on all three evaluation metrics. The ablation study confirms that each of the three novel mechanisms contributes independently and significantly to BAQIO's overall performance.

### Author Contributions

**O.H**: Conceptualization, Methodology, Software, Formal analysis, writing — original draft, Writing — review and editing, Project administration. **E.L**: Methodology, Software, Validation, Data curation, Visualization. **A.A**: Investigation, Formal analysis, EEG preprocessing pipeline, Writing — review and editing. **A.M**: Supervision, Resources, Writing — review and editing, Funding acquisition.

### Funding

This research received no external funding.

### Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data Availability

All 18 UCI datasets are freely accessible at https://archive.ics.uci.edu/ml/.

## References

[1] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.

[2] Y. Zhang, S. Wang, P. Phillips, G. Ji, Binary PSO with mutation operator for feature selection using decision tree applied to spam detection, Knowl.-Based Syst. 64 (2014) 22–31.

[3] A. Selvakumar, K. Muneeswaran, Firefly algorithm based feature selection for network intrusion detection, Comput. Secur. 81 (2019) 148–155.

[4] M. Mafarja, S. Mirjalili, Whale optimization approaches for wrapper feature selection, Appl. Soft Comput. 62 (2018) 441–453.

[5] M. Dash, H. Liu, Feature selection for classification, Intell. Data Anal. 1 (1997) 131–156.

[6] H. Liu, L. Yu, Toward integrating feature selection algorithms for classification and clustering, IEEE Trans. Knowl. Data Eng. 17 (2005) 491–502.

[7] S. Mirjalili, A.H. Gandomi, Chaotic gravitational constants for the gravitational search algorithm, Appl. Soft Comput. 53 (2017) 407–419.

[8] E. Hancer, B. Xue, M. Zhang, Differential evolution for filter feature selection based on information theory and feature ranking, Knowl.-Based Syst. 140 (2018) 103–119.

[9] R. Eberhart, J. Kennedy, A new optimizer using particle swarm theory, in: Proc. MHSPC, 1995, pp. 39–43.

[10] S. Emary, H.M. Zawbaa, A.E. Hassanien, Binary grey wolf optimization approaches for feature selection, Neurocomputing 172 (2016) 371–381.

[11] M. Mafarja, S. Mirjalili, Whale optimization approaches for wrapper feature selection, Appl. Soft Comput. 62 (2018) 441–453.

[12] M.A. Elaziz, D. Yousri, S. Mirjalili, A hybrid Harris hawks-moth flame optimization algorithm including fractional-order chaos maps and evolutionary population dynamics, Adv. Eng. Softw. 154 (2021) 102973.

[13] A.A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, H. Chen, Harris hawks optimization: Algorithm and applications, Future Gener. Comput. Syst. 97 (2019) 849–872.

[14] M. Mafarja, S. Mirjalili, Efficient hybrid nature-inspired binary optimizers for feature selection, Cogn. Comput. 12 (2020) 150–175.

[15] B. Crawford, R. Soto, M. Astorga, J. García, C. Castro, F. Paredes, Putting continuous metaheuristics to work in binary search spaces, Complexity 2017 (2017) 1–19.

[16] S. Rahnamayan, H.R. Tizhoosh, M.M.A. Salama, Opposition-based differential evolution, IEEE Trans. Evol. Comput. 12 (2008) 64–79.

[17] K.-H. Han, J.-H. Kim, Quantum-inspired evolutionary algorithm for a class of combinatorial optimization, IEEE Trans. Evol. Comput. 6 (2002) 580–593.

[18] Y. Zhang, S. Wang, G. Ji, A comprehensive survey on particle swarm optimization algorithm and its applications, Math. Probl. Eng. 2015 (2015) 1–38.

[19] X. Li, J. Zhang, M. Yin, Animal migration optimization: An optimization algorithm inspired by animal migration behavior, Neural Comput. Appl. 24 (2014) 1867–1877.

[20] P. Larrañaga, J.A. Lozano (Eds.), Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation, Kluwer Academic Publishers, Boston, 2002.

[21] C.R. Raquel, P.C. Naval, An effective use of crowding distance in multiobjective particle swarm optimization, in: Proc. GECCO 2005, ACM, 2005, pp. 257–264.

[22] R. Kohavi, G.H. John, Wrappers for feature subset selection, Artif. Intell. 97 (1997) 273–324.

[23] E.-G. Talbi, Metaheuristics: From Design to Implementation, Wiley, Hoboken, NJ, 2009.

[24] S. Mirjalili, SCA: A sine cosine algorithm for solving optimization problems, Knowl.-Based Syst. 96 (2016) 120–133.

[25] L. Abualigah, M. Abd Elaziz, P. Sumari, Z.W. Geem, A.H. Gandomi, Reptile Search Algorithm (RSA): A nature-inspired meta-heuristic optimizer, Expert Syst. Appl. 191 (2022) 116158.

[26] Z. Beheshti, BMPA-V-shaped: A binary marine predators algorithm using V-shaped transfer function for feature selection, J. Ambient Intell. Humaniz. Comput. 14 (2023) 1–21.

[27] H. Wang, Z. Wu, S. Rahnamayan, Y. Liu, M. Ventresca, Enhancing PSO using generalized opposition-based learning, Inf. Sci. 181 (2011) 4699–4714.

[28] R.N. Mantegna, Fast, accurate algorithm for numerical simulation of Lévy stable stochastic processes, Phys. Rev. E 49 (1994) 4677–4683.

[29] D. Dua, C. Graff, UCI Machine Learning Repository, University of California, Irvine, CA, 2019. http://archive.ics.uci.edu/ml.

[30] M. Mafarja, I. Aljarah, A.A. Heidari, A.I. Hammouri, H. Faris, A.-Z. Ala'M, S. Mirjalili, Evolutionary population dynamics and grasshopper optimization approaches for feature selection problems, Knowl.-Based Syst. 145 (2018) 25–45.

[31] A.E. Hassanien, I. Yamany, H.M. Zawbaa, Binary dragonfly algorithm for feature selection, in: Proc. AECIA 2017, Springer, 2017, pp. 309–318.

[32] Q. Al-Tashi, S.J. Abdulkadir, H.M. Rais, S. Mirjalili, H. Alhussian, Approaches to multi-objective feature selection: A systematic literature review, IEEE Access 8 (2020) 125076–125096.

[33] N.H. Awad, M.Z. Ali, J.J. Liang, B.Y. Qu, P.N. Suganthan, Problem Definitions and Evaluation Criteria for the CEC 2017 Special Session on Numerical Optimization, Technical Report, Nanyang Technological University, 2016.

[34] A. Tharwat, T. Gaber, Classification based on genetic programming approach: A comprehensive survey, IEEE Access 9 (2021) 146064–146075.

[35] Z. Dong, S. Wang, N. Ahmad, I. Ullah, S. Waseem, Y. Zheng, Enhanced binary dragonfly algorithm for feature selection in high-dimensional data, Expert Syst. Appl. 227 (2023) 120294.

[36] I. Aljarah, A.M. Al-Zoubi, H. Faris, M.A. Hassonah, S. Mirjalili, H. Saadeh, Simultaneous feature selection and support vector machine optimization using the grasshopper optimization algorithm, Cogn. Comput. 10 (2018) 478–495.

[37] J. Too, A.R. Abdullah, N. Mohd Saad, W. Tee, EMG feature selection and classification using a Pbest-guide binary particle swarm optimization, Computation 7 (2019) 12.

[38] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, Adv. Eng. Softw. 69 (2014) 46–61.

[39] S. Mirjalili, A. Lewis, The whale optimization algorithm, Adv. Eng. Softw. 95 (2016) 51–67.

[40] S. Mirjalili, A.H. Gandomi, S.Z. Mirjalili, S. Saremi, H. Faris, S.M. Mirjalili, Salp swarm algorithm: A bio-inspired optimizer for engineering design problems, Adv. Eng. Softw. 114 (2017) 163–191.